FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		ATTORNEY'S DOCKET NUMBER **PF-0661 USN** U.S. APPLICATION NO. (If known, see 37 CFR 1.5) **TO BE ASSIGNED** 09/889617
INTERNATIONAL APPLICATION NO. **PCT/US00/01565**	INTERNATIONAL FILING DATE **21 January 2000**	PRIORITY DATE CLAIMED **22 January 1999**

TITLE OF INVENTION
CANCER-ASSOCIATED PROTEINS

APPLICANT(S) FOR DO/EO/US
INCYTE PHARMACEUTICALS, INC.; HILLMAN, Jennifer L.; YUE, Henry; TANG, Y. Tom; AZIMZAI, Yalda

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: **EL 856 154 101 US**
4) Request to Transfer

U.S. APPLICATION NO. (If known, see 37 CFR 1.5) TO BE ASSIGNED 09/889617	INTERNATIONAL APPLICATION NO.: PCT/US00/01565	ATTORNEY'S DOCKET NUMBER PF-0661 USN

17. ☒ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.....$1000.00

☐International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..$860.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$710.00

☒**International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4).....$690.00**

☐International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4).....$100.00

ENTER APPROPRIATE BASIC FEE AMOUNT =				$690.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	23 =	3	X **$ 18.00**	$ 54.00	
Independent Claims	2 =	0	X **$ 80.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$744.00	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
SUBTOTAL =				$744.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest clailmed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$744.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$	
TOTAL FEES ENCLOSED =				$744.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $_____ to cover the above fees is enclosed.

b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $ 744.00 to cover the above fees.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 16 July 2001

Docket No.: PF-0661 USN

09/889617

JC18 Rec'd PCT/PTO 16 JUL 2001

"Express Mail" mailing label number EL 856 154 101 US. I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under37 CFR § 1.10, addressed to: Commissioner for Patents, Box Provisional Patent Application, Washington, D.C. 20231 on 17 July 2001.

By: Nancy Ramos Printed: NANCY RAMOS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Jennifer L. Hillman, Henry Yue, Y. Tom Tang & Yalda Azimzai		
Title:	CANCER-ASSOCIATED PROTEINS		
Serial No.:	To Be Assigned	Filed:	Herewith
Examiner:	To Be Assigned	Group Art Unit:	To Be Assigned

Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

REQUEST TO TRANSFER

Sir:

With regard to the requirement of 37 CFR 1.821(e) which requires that a copy of the Sequence Listing in computer readable form (CRF) be submitted, Applicants state that the paper copy of the Sequence Listing for the instant application is identical with the computer readable form filed with PCT Application No. PCT/US00/01565, filed 21 January 2000, in the US/RO to which priority is claimed. In accordance with 37 C.F.R. §1.821(e), please use the computer readable form filed with PCT Application No. PCT/US00/01565 as the computer readable form for the instant application. It is understood that the Patent and Trademark Office will make the necessary change in application number and filing date for the computer readable form that will be used for the instant application.

Applicants believe that no fee is due with this communication. However, if the USPTO determines that a fee is due, the Commissioner is hereby authorized to charge Deposit Account No. **09-0108. This form is enclosed in duplicate.**

Respectfully submitted,

INCYTE GENOMICS, INC.

Date: 16 July 2001

Diana Hamlet-Cox
Reg. No. 33,302
Direct Dial Telephone: (650) 845-4639

3160 Porter Drive
Palo Alto, California, 94304
Tel. No. 650-855-0555
Fax. No. 650-849-88860

CANCER-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cancer-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and autoimmune/inflammatory disorders.

BACKGROUND OF THE INVENTION

Cancers, also called neoplasias, are characterized by continuous cell proliferation. They can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant new growth soft epithelial cells that may infiltrate surrounding tissues and give rise to metastases. Sarcomas may be of epithelial origin or arise from connective tissue. Leukemias are progressive malignancies of blood-forming tissue characterized by development of leukocytes and their precursors, and may be classified as myelogenous (bone marrow) or lymphocytic. Tumor cells vary; malignant cells may be quite similar to normal cells within the tissue of origin or may be undifferentiated (anaplastic). Tumor cells may posses few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contain large areas of ischemic necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin. Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or through hematogenous spread.

Cancers are associated with oncoproteins which are capable of transforming normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein while others are abnormally expressed with respect to location or level of expression. Normal cell proliferation begins with binding of a growth factor to its receptor on the cell membrane, resulting in activation of a signal system that induces and activates nuclear regulatory factors to initiate DNA transcription, subsequently leading to cell division. Classes of oncoproteins known to affect the cell cycle controls include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Generally, oncoproteins lack normal regulatory elements and are expressed independently of growth factors or other external signals such as cell adhesion.

Oncoproteins are encoded by genes, called oncogenes, that are derived from genes that normally control cell growth and development. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset

of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22. The hybrid c-abl-bcr gene encodes a chimeric protein that has tyrosine kinase activity. In chronic myeloid leukemia, the chimeric protein has a molecular weight of 210 kd, whereas in acute leukemias a more active 180 kd tyrosine kinase is formed (Robbins, S.L. et al. (1994) Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA).

Immunological defenses against cancer include induction of apoptosis in mutant cells by tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Examples of tumor antigens are those encoded by the GAGE genes which are expressed in a variety of tumors and in the testis. GAGE-7 is expressed specifically in testicular and placental tissue and has been associated with prostate cancer progression from its earlier androgen-sensitive stage to the later androgen-insensitive stage (Chen, M.E. et al. (1998) J. Biol. Chem. 273:17618-17625).

Some cancers involve suppression of normal apoptotic cell death. In another example of chromosomal translocation, the E2A-HLF fusion associated with pro-B acute leukemia links transactivation domains of the E2A gene to the bZIP DNA binding and protein dimerization region of HLF, a leucine zipper-type transcription factor normally expressed in liver, kidney, and CNS neuronal cells. The chimeric protein appears to alter the transcriptional control of genes responsible for apoptosis in pro-B cells so defective cells that would normally self-destruct survive and proliferate. One target of the E2A-HLF gene protein is SRPUL (sushi-repeat protein upgraded in leukemia). SRPUL contains consensus repeat motifs that appear to function in cell adhesion (Kurosawa, H. et al. (1999) Blood 93:321-332).

The discovery of new cancer-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative and autoimmune/inflammatory disorders.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, cancer-associated proteins, referred to collectively as "CAP" and individually as "CAP-1," "CAP-2," and "CAP-3." In one aspect, the invention provides an isolated polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. In one alternative, the invention provides an isolated

polypeptide comprising the amino acid sequence of SEQ ID NO:1-3.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. In one alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:4-6.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3.

The invention further provides an isolated polynucleotide comprising a) a polynucleotide

sequence selected from the group consisting of SEQ ID NO:4-6, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:4-6, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:4-6, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:4-6, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). The method comprises a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, and a pharmaceutically acceptable excipient. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. The method comprises a) exposing a sample comprising the polypeptide to a compound,

and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-3. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:4-6, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figures 1A, 1B, 1C, 1D, 1E, and 1F show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:4) of CAP-1. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, 2C, 2D, and 2E show the amino acid sequence (SEQ ID NO:2) and nucleic acid sequence (SEQ ID NO:5) of CAP-2. The alignment was produced using MACDNASIS PRO software.

Figures 3A and 3B show the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:6) of CAP-3. The alignment was produced using MACDNASIS PRO software.

Figures 4A and 4B show the amino acid sequence alignment between CAP-1 (Incyte Clone 1518859; SEQ ID NO:1) and rat drs (downregulated by v-src) gene product (GI 1345423; SEQ ID

NO:7), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 5A, 5B, and 5C show the amino acid sequence alignment between CAP-2 (Incyte Clone 2616269; SEQ ID NO:2) and human BCR gene protein (GI 487348; SEQ ID NO:8), produced using the multisequence alignment program of LASERGENE software.

Figure 6 shows the amino acid sequence alignment between CAP-3 (Incyte Clone 3117642; SEQ ID NO:3) and human GAGE-7 prostate tumor antigen (GI 3511023; SEQ ID NO:9), produced using the multisequence alignment program of LASERGENE software.

Table 1 shows the tools, programs, and algorithms used to analyze CAP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CAP" refers to the amino acid sequences of substantially purified CAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CAP either by directly interacting with CAP or by acting on components of the biological pathway in which CAP participates.

An "allelic variant" is an alternative form of the gene encoding CAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CAP or a polypeptide with at least one functional characteristic of CAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.

Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CAP either by directly interacting with CAP or by acting on components of the biological pathway in which CAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules

may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CAP or fragments of CAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu

Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of CAP or the polynucleotide encoding CAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:4-6 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:4-6, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:4-6 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:4-6 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:4-6 and the region of SEQ ID NO:4-6 to

which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-3 is encoded by a fragment of SEQ ID NO:4-6. A fragment of SEQ ID NO:1-3 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-3. For example, a fragment of SEQ ID NO:1-3 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-3. The precise length of a fragment of SEQ ID NO:1-3 and the region of SEQ ID NO:1-3 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS

8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of CAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding CAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are

isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing

selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CAP, or fragments thereof, or CAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of

the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human cancer-associated proteins (CAP), the polynucleotides encoding CAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and autoimmune/inflammatory disorders.

Nucleic acids encoding the CAP-1 of the present invention were identified in Incyte Clone 1518859H1 from the human bladder tumor cDNA library (BLADTUT04) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was assembled from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1518859H1 (BLADTUT04), 2634789F6 (COLNTUT15), 2762312F6 (BRSTNOT12) and sequences SBCA02155F1, SBCA05656F1, SBIA02789D1, and SBIA03101D1 (derived from pooled cDNA libraries).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C, 1D, 1E, and 1F. CAP-1 is 465 amino acids in length and has a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S112; 11 potential casein kinase II phosphorylation sites at residues S36, T85, S143, S170, T242, S279, S348, T377, T436, T442, and S450; 10 potential protein kinase C phosphorylation sites at residues T6, S73, S108, T200, T222, S252, T266, S308, S353, and T455; and a potential tyrosine kinase phosphorylation site at residue Y431. PFAM analysis shows that CAP-1 shares homology with a Sushi (short consensus repeat) domain motif from residues C59 to C117, C122 to C176, and C264 to C319. SPScan and HMM analyses show that CAP-1 has a potential signal peptide from residue M1 to about T22. As shown in Figures 4A and 4B, CAP-1 has chemical and structural similarity with rat drs (downregulated by v-src) gene product (GI 1345423; SEQ ID NO:7). In particular, CAP-1 and rat drs gene product share 44% identity. A fragment of SEQ ID NO:4 from about nucleotide 1057 to about nucleotide 1101 is useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:4 and to distinguish between SEQ ID NO:4 and a related sequence. The encoded polypeptide is useful, for example, as an immunogenic peptide. Northern analysis shows the expression of this sequence in various libraries, at least 57% of which are associated with cell proliferation and at least 24% of which are associated with inflammation and the immune response. Of particular note is the expression of CAP-1 in cardiovascular (33%), reproductive (29%), and gastrointestinal tissues (14%).

Nucleic acids encoding the CAP-2 of the present invention were identified in Incyte Clone 2616269H1 from the human gall bladder cDNA library (GBLANOT01) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:5, was assembled from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2616269H1 (GBLANOT01), 3557735H1 (LUNGNOT31), 1965572R6 (BRSTNOT04), 2182972F6 (SININOT01), 1857443F6 (PROSNOT18), 3568062H1 (HEAPNOT01), 1878060T6 (LEUKNOT03), 1795056R6 (PROSTUT05), and 1486038H1 (CORPNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2, as shown in Figures 2A, 2B, 2C, 2D, and 2E. CAP-2 is 278 amino acids in length and has a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue T50; 3 potential casein kinase II phosphorylation sites at residues S29, T50, and S187; and 5 potential protein kinase C phosphorylation sites at residues T183, T211, S219, S312, and T380. PFAM analysis shows that CAP-2 shares homology with a WD domain G-beta repeat motif from residues R141 to K177, S180 to P216, G223 to D264, and M271 to D308. PRINTS analysis shows that CAP-2 shares homology with a WD domain G-beta repeat motif from residues I164 to I178, I203 to L217, and L295 to L309. SPScan analysis shows that CAP-2 has a potential signal peptide from residue M1 to about A35. As shown in Figures 5A, 5B, and 5C, CAP-2 has chemical and structural similarity with the human BCR gene protein (GI 487348; SEQ ID NO:8). In particular, CAP-2 and the human bcr-abl gene protein share 44% identity. A fragment of SEQ ID NO:5 from about nucleotide 483 to about nucleotide 527 is useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:5 and to distinguish between SEQ ID NO:5 and a related sequence. The encoded polypeptide is useful, for example, as an immunogenic peptide. Northern analysis shows the expression of this sequence in various libraries, at least 52% of which are associated with cell proliferation and at least 34% of which are associated with inflammation and the immune response. Of particular note is the expression of CAP-2 in reproductive (30%), nervous (22%), and gastrointestinal tissues (16%).

Nucleic acids encoding the CAP-3 of the present invention were identified in Incyte Clone 3117642H1 from the human lung tumor cDNA library (LUNGTUT13) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:6, was assembled from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 3117642H1 and 3117642F6 (LUNGTUT13), 1980062H1 (LUNGTUT03), and 791000R1 (PROSTUT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3, as shown in Figures 3A and 3B. CAP-3 is 400 amino acids in length and

has 6 potential casein kinase II phosphorylation sites at residues S19, S23, T41, T71, T113, and T136; and a potential protein kinase C phosphorylation site at residue T81. As shown in Figure 6, CAP-2 has chemical and structural similarity with human GAGE-7 prostate tumor antigen (GI 3511023; SEQ ID NO:9). In particular, CAP-3 and human GAGE-7 prostate tumor antigen share 18% identity. A fragment of SEQ ID NO:6 from about nucleotide 117 to about nucleotide 159, and a fragment of SEQ ID NO:6 from about nucleotide 262 to about nucleotide 306, are each useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:6 and to distinguish between SEQ ID NO:6 and a related sequence. The polypeptide encoded by each fragment is useful, for example, as an immunogenic peptide. Northern analysis shows the expression of this sequence in four libraries, all of which are associated with cell proliferation. Two of the libraries are from lung tumor tissue, one is from prostate tumor tissue, and one is from a K562 chronic myelogenous leukemia precursor cell line.

The invention also encompasses CAP variants. A preferred CAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CAP amino acid sequence, and which contains at least one functional or structural characteristic of CAP.

The invention also encompasses polynucleotides which encode CAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:4-6, which encodes CAP.

The invention also encompasses a variant of a polynucleotide sequence encoding CAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:4-6 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:4-6. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the

polynucleotide sequence of naturally occurring CAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CAP and CAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:4-6 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a

variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CAP may be cloned in recombinant DNA molecules that direct expression of CAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids

Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of CAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active CAP, the nucleotide sequences encoding CAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CAP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CAP. Transcription of sequences encoding CAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology

(1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CAP in cell lines is preferred. For example, sequences encoding CAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins

(GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CAP is inserted within a marker gene sequence, transformed cells containing sequences encoding CAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CAP and that express CAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety

of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CAP may be designed to contain signal sequences which direct secretion of CAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CAP encoding sequence and the heterologous protein sequence, so that CAP may be cleaved away from the heterologous moiety following purification.

Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of CAP may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of CAP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CAP and cancer-associated proteins. In addition, the expression of CAP is closely associated with cell proliferative and autoimmune/inflammatory disorders. Therefore, CAP appears to play a role in cell proliferative and autoimmune/inflammatory disorders. In the treatment of disorders associated with increased CAP expression or activity, it is desirable to decrease the expression or activity of CAP. In the treatment of disorders associated with decreased CAP expression or activity, it is desirable to increase the expression or activity of CAP.

Therefore, in one embodiment, CAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CAP. Examples of such disorders include, but are not limited to, a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal

dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing CAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CAP. Examples of such disorders include, but are not limited to, those cell proliferative and autoimmune/inflammatory disorders described above. In one aspect, an antibody which specifically binds CAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the

various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CAP may be produced using methods which are generally known in the art. In particular, purified CAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CAP. Antibodies to CAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single

chain antibodies may be adapted, using methods known in the art, to produce CAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CAP epitopes, represents the average affinity, or avidity, of the antibodies for CAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell,

J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CAP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CAP. Thus, complementary molecules or fragments may be used to modulate CAP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding CAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CAP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing

is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable

for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CAP, antibodies to CAP, and mimetics, agonists, antagonists, or inhibitors of CAP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose,

hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any

or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CAP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CAP or fragments thereof, antibodies of CAP, and agonists, antagonists or inhibitors of CAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of

about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CAP may be used for the diagnosis of disorders characterized by expression of CAP, or in assays to monitor patients being treated with CAP or agonists, antagonists, or inhibitors of CAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CAP include methods which utilize the antibody and a label to detect CAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CAP expression. Normal or standard values for CAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to CAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CAP, and to monitor regulation of CAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CAP or closely related molecules may be used to identify nucleic acid sequences which encode CAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CAP, allelic variants, or related

sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:4-6 or from genomic sequences including promoters, enhancers, and introns of the CAP gene.

Means for producing specific hybridization probes for DNAs encoding CAP include the cloning of polynucleotide sequences encoding CAP or CAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CAP may be used for the diagnosis of disorders associated with expression of CAP. Examples of such disorders include, but are not limited to, a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The

polynucleotide sequences encoding CAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of CAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra,

pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CAP, or fragments thereof, and washed. Bound CAP is then detected by methods well known in the art. Purified CAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CAP specifically compete with a test compound for binding CAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CAP.

In additional embodiments, the nucleotide sequences which encode CAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0661 P, filed January 22, 1999], are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

CAP-1

The BLADTUT04 library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Family history included type I diabetes, a malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a Polytron PT 3000 homogenizer (Brinkmann Instruments, Westbury NY). RNA was isolated as per Stratagene's RNA isolation protocol (Stratagene, La Jolla CA). RNA was extracted twice with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase. Poly (A+) RNA was isolated using the OLIGOTEX kit (QIAGEN, Chatsworth CA).

CAP-2

The GBLANOT01 library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.

The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml TRIZOL; Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate, using a Polytron PT 3000 homogenizer (Brinkmann Instruments). After brief incubation on ice, chloroform was added

(1:5 v/v), and the mixture was centrifuged to separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and treated with DNase. The RNA was re-extracted with acid phenol-chloroform and reprecipitated with sodium acetate and ethanol. Poly(A+) RNA was isolated using the OLIGOTEX mRNA purification kit (QIAGEN).

CAP-3

The LUNGTUT13 library was constructed using RNA isolated from tumorous lung tissue removed from the right upper lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology indicated invasive grade 3 (of 4) adenocarcinoma. Family history included atherosclerotic coronary artery disease, and type II diabetes.

The frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml TRIZOL; Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate, using a Polytron PT 3000 homogenizer (Brinkmann Instruments). After brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged to separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and treated with DNase. The RNA was re-extracted with acid phenol-chloroform and reprecipitated with sodium acetate and ethanol. Poly(A+) RNA was isolated using the OLIGOTEX mRNA purification kit (QIAGEN).

Poly(A+) RNA was used for cDNA synthesis and construction of all three cDNA libraries according to the recommended protocols in the SUPERSCRIPT plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech), and those cDNAs exceeding 400 bp were ligated into pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into DH5α competent cells or ElectroMAX®cells (Life Technologies).

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 1 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST,

dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:4-6. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which

the transcript encoding CAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in the description of the invention.

V. Extension of CAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:4-6 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:4-6 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:4-6 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7

counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the CAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO

4.06 software (National Biosciences) and the coding sequence of CAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CAP-encoding transcript.

IX. Expression of CAP

Expression and purification of CAP is achieved using bacterial or virus-based expression systems. For expression of CAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CAP obtained by these methods can be used directly in the following activity assay.

X. Functional Assays

CAP function is assessed by expressing the sequences encoding CAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XI. Production of CAP Specific Antibodies

CAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CAP amino acid sequence is analyzed using LASERGENE software

(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CAP activity by, for example, binding the peptide or CAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XII. Purification of Naturally Occurring CAP Using Specific Antibodies

Naturally occurring or recombinant CAP is substantially purified by immunoaffinity chromatography using antibodies specific for CAP. An immunoaffinity column is constructed by covalently coupling anti-CAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CAP is collected.

XIII. Identification of Molecules Which Interact with CAP

CAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CAP, washed, and any wells with labeled CAP complex are assayed. Data obtained using different concentrations of CAP are used to calculate values for the number, affinity, and association of CAP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 1 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:2-3,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:2-3,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:2-3, or

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:2-3.

2. An isolated polypeptide of claim 1, having an amino acid sequence selected from the group consisting of SEQ ID NO:2-3.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide of claim 3, having a sequence selected from the group consisting of SEQ ID NO:5-6.

5. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

6. A cell transformed with a recombinant polynucleotide of claim 5.

7. A transgenic organism comprising a polynucleotide of claim 5.

8. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

9. An isolated antibody which specifically binds to a polypeptide of claim 1.

10. An isolated polynucleotide comprising:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-6,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-6,

c) a polynucleotide sequence complementary to a), or

d) a polynucleotide sequence complementary to b).

11. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 10.

12. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 10, the method comprising:

a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

13. A method of claim 12, wherein the probe comprises at least 30 contiguous nucleotides.

14. A method of claim 12, wherein the probe comprises at least 60 contiguous nucleotides.

15. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

16. A method of treating a disease or condition associated with decreased expression of functional CAP, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 15.

17. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

18. A pharmaceutical composition comprising an agonist compound identified by a method of claim 17 and a pharmaceutically acceptable excipient.

19. A method of treating a disease or condition associated with decreased expression of functional CAP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 18.

20. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

21. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with overexpression of functional CAP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 21.

23. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 4, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/10, 5/10, C07K 14/47, 16/18, A01K 67/027, A61K 38/17, G01N 33/50, C12Q 1/68

A2

(11) International Publication Number: **WO 00/43508**

(43) International Publication Date: 27 July 2000 (27.07.00)

(21) International Application Number: PCT/US00/01565

(22) International Filing Date: 21 January 2000 (21.01.00)

(30) Priority Data:
60/183,027 22 January 1999 (22.01.99) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/236,205 (CIP)
Filed on 22 January 1999 (22.01.99)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US).

(74) Agent: HAMLET-COX, Diana; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CANCER-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human cancer-associated proteins (CAP) and polynucleotides which identify and encode CAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CAP.

```
                9          18          27          36          45          54
5' CTG CAG CAG ACG GAC TGA GTT CCT CTA ATC CCT GTG TTC CTT CTC CCC CAT CTT

               63          72          81          90          99         108
   TCT AAA ACC CTT CTC TGA GAG AGG AAT AAC TAT AGC TTC AGG GAT AAT ATA GCT

              117         126         135         144         153         162
   TTA AGG AAA CTT TTG GCA GAT GTG GAC GTC GTA ACA TCT GGG CAG TGT TAA CAG

              171         180         189         198         207         216
   AAT CCC GGA GGC CGG GAC AGA CCA GGA GCC ACT CGT TCT AGG AAT GTT AAA GTA

              225         234         243         252         261         270
   GAA GGT TTT TTC CAA TTG ATG AGA GGA GCA GAG AGG AAG GAG AAA GAG GAG GAG

              279         288         297         306         315         324
   AGA GAA AAA GGG CAC AAA ATA CCA TAA AAC AGA TCC CAT ATT TCT GCT TCC CCT

              333         342         351         360         369         378
   CAC TTT TAG AAG TTA ATT GAT GGC TGA CTT CTG AAA GTC ACT TTC CTT TGC CCT

              387         396         405         414         423         432
   GGT ACT TCA GGC CAT ATA CAT CTT TTC TTG TCT CCA TAA TCC TCC CTT TCA AGG

              441         450         459         468         477         486
   ATG GCC AGT CAG CTA ACT CAA AGA GGA GCT CTC TTT CTG CTG TTC TTC CTA ACT
   M   A   S   Q   L   T   Q   R   G   A   L   F   L   L   F   F   L   T
```

FIGURE 1A

```
                495                 504                 513                 522                 531                 540
CCG GCA GTG ACA CCA ACA TGG TAT GCA GGT TCT GGC TAC TAT CCG GAT GAA AGC
P   A   V   T   P   T   W   Y   A   G   S   G   Y   Y   P   D   E   S

                549                 558                 567                 576                 585                 594
TAC AAT GAA GTA TAT GCA GAG GAG GTC CCA CAG GCT CCT GCC CTG GAC TAC CGA
Y   N   E   V   Y   A   E   E   V   P   Q   A   P   A   L   D   Y   R

                603                 612                 621                 630                 639                 648
GTC CCC CGA TGG TGT TAT ACA TTA AAT ATC CAG GAT GGA GAA GCC ACA TGC TAC
V   P   R   W   C   Y   T   L   N   I   Q   D   G   E   A   T   C   Y

                657                 666                 675                 684                 693                 702
TCA CCG AAG GGA GGA AAT TAT CAC AGC AGC CTG GGC ACG CGT TGT GAG CTC TCC
S   P   K   G   G   N   Y   H   S   S   L   G   T   R   C   E   L   S

                711                 720                 729                 738                 747                 756
TGT GAC CGG GGC TTT CGA TTG ATT GGA AGG AGG TCG GTG CAA TGC CTG CCA AGC
C   D   R   G   F   R   L   I   G   R   R   S   V   Q   C   L   P   S

                765                 774                 783                 792                 801                 810
CGT CGT TGG TCT GGA ACT GCC TAC TGC AGG CAG ATG AGA TGC CAC GCA CTA CCA
R   R   W   S   G   T   A   Y   C   R   Q   M   R   C   H   A   L   P

                819                 828                 837                 846                 855                 864
TTC ATC ACT AGT GGC ACT TAC ACC TGC ACA AAT GGA GTG CTT CTT GAC TCT CGC
F   I   T   S   G   T   Y   T   C   T   N   G   V   L   L   D   S   R
```

FIGURE 1B

```
          873         882         891         900         909         918
TGT GAC TAC AGC TGT TCC AGT GGC TAC CAC CTG GAA GGT GAT CGC AGC CGA ATC
C   D   Y   S   C   S   S   G   Y   H   L   E   G   D   R   S   R   I

          927         936         945         954         963         972
TGC ATG GAA GAT GGG AGA TGG AGT GGA GGC GAG CCT GTA TGT GTA GAC ATA GAT
C   M   E   D   G   R   W   S   G   G   E   P   V   C   V   D   I   D

          981         990         999        1008        1017        1026
CCC CCC AAG ATC CGC TGT CCC CAC TCA CGT GAG AAG ATG GCA GAG CCA GAG AAA
P   P   K   I   R   C   P   H   S   R   E   K   M   A   E   P   E   K

         1035        1044        1053        1062        1071        1080
TTG ACT GCT CGA GTA TAC TGG GAC CCA CCG TTG GTG AAA GAT TCT GCT GAT GGT
L   T   A   R   V   Y   W   D   P   P   L   V   K   D   S   A   D   G

         1089        1098        1107        1116        1125        1134
ACC ATC ACC AGG GTG ACA CTT CGG GGC CCT GAG CCT GGC TCT CAC TTT CCC GAA
T   I   T   R   V   T   L   R   G   P   E   P   G   S   H   F   P   E

         1143        1152        1161        1170        1179        1188
GGA GAG CAT GTG ATT CGT TAC ACT GCC TAT GAC CGA GCC TAC AAC CGG GCC AGC
G   E   H   V   I   R   Y   T   A   Y   D   R   A   Y   N   R   A   S

         1197        1206        1215        1224        1233        1242
TGC AAG TTC ATT GTG AAA GTA CAA GTG AGA CGC TGC CCA ACT CTG AAA CCT CCG
C   K   F   I   V   K   V   Q   V   R   R   C   P   T   L   K   P   P
```

FIGURE 1C

```
      1251        1260        1269        1278        1287        1296
CAG CAC GGC TAC CTC ACC TGC ACC TCA GCG GGG GAC AAC TAT GGT GCC ACC TGT
Q   H   G   Y   L   T   C   T   S   A   G   D   N   Y   G   A   T   C

      1305        1314        1323        1332        1341        1350
GAA TAC CAC TGT GAT GGC GGT TAT GAT CGC CAG GGG ACA CCC TCC CGG GTC TGT
E   Y   H   C   D   G   G   Y   D   R   Q   G   T   P   S   R   V   C

      1359        1368        1377        1386        1395        1404
CAG TCC AGC CGC CAG TGG TCA GGT TCA CCA CCA ATC TGT GCT CCT ATG AAG ATT
Q   S   S   R   Q   W   S   G   S   P   P   I   C   A   P   M   K   I

      1413        1422        1431        1440        1449        1458
AAC GTC AAC GTC AAC TCA GCT GCT GGT CTC TTG GAT CAA TTC TAT GAG AAA CAG
N   V   N   V   N   S   A   A   G   L   L   D   Q   F   Y   E   K   Q

      1467        1476        1485        1494        1503        1512
CGA CTC CTC ATC ATC TCA GCT CCT GAT CCT TCC AAC CGA TAT TAT AAA ATG CAG
R   L   L   I   I   S   A   P   D   P   S   N   R   Y   Y   K   M   Q

      1521        1530        1539        1548        1557        1566
ATC TCT ATG CTA CAG CAA TCC ACC TGT GGA CTG GAT TTG CGG CAT GTG ACC ATC
I   S   M   L   Q   Q   S   T   C   G   L   D   L   R   H   V   T   I

      1575        1584        1593        1602        1611        1620
ATT GAA CTG GTG GGA CAG CCA CCT CAG GAG GTG GGG CGC ATC CGG GAG CAA CAG
I   E   L   V   G   Q   P   P   Q   E   V   G   R   I   R   E   Q   Q
```

FIGURE 1D

```
                 1629           1638           1647           1656           1665           1674
CTG TCA GCC AAC ATC ATC GAG GAG CTC AGG CAA TTT CAG CGC CTC ACT CGC TCC
L   S   A   N   I   I   E   E   L   R   Q   F   Q   R   L   T   R   S

                 1683           1692           1701           1710           1719           1728
TAC TTC AAC ATG GTG TTG ATT GAC AAG CAG GGT ATT GAC CGA GAC CGC TAC ATG
Y   F   N   M   V   L   I   D   K   Q   G   I   D   R   D   R   Y   M

                 1737           1746           1755           1764           1773           1782
GAA CCT GTC ACC CCC GAG GAA ATC TTC ACA TTC ATT GAT GAC TAC CTA CTG AGC
E   P   V   T   P   E   E   I   F   T   F   I   D   D   Y   L   L   S

                 1791           1800           1809           1818           1827           1836
AAT CAG GAG TTG ACC CAG CGT CGG GAG CAA AGG GAC ATA TGC GAG TGA ACT TGA
N   Q   E   L   T   Q   R   R   E   Q   R   D   I   C   E

                 1845           1854           1863           1872           1881           1890
GCC AGG GCA TGG TTA AAG TCA AGG GAA AAG CTC CTC TAG TTA GCT GAA ACT GGG

                 1899           1908           1917           1926           1935           1944
ACC TAA TAA AAG GAG GAA ATG TTT TCC CAC AGT TCT AGG GAC AGG ACT CTG AGG

                 1953           1962           1971           1980           1989           1998
TGG GTG AGT TTG ACA AAT CCT GCA GTG TTT CCA GGC ATC CTT TTA GGA CTG TGT

                 2007           2016           2025           2034           2043           2052
AAT AGT TTC CCT AGA AGC TAG GTA GGG ACT GAG GAC AGG CCT TGG GCA GTG GGT
```

FIGURE 1E

```
2061       2070       2079       2088       2097       2106
TGG GGG TAG AAG TTC TTC CTT TCC TAA CCC GGG CCC CTG CCC AGC TCT CCA AAG

2115       2124       2133       2142       2151
TCT TTC AGA AAA GTA AAT CCT AAA TTC AGT GAA AAA AAA AAA AAA A 3'
```

FIGURE 1F

```
                11          20          29          38          47          56
5'CACGC GGT CAG GCT TGG GCC GAC ATC GCG GGG ACA GGG GTG GCC ATG GCG GCT CGG
                                                        M   A   A   R

          65          74          83          92          101         110
   GAG TCG GCT GCC CGC CCG GCC GCG GGG CCT GCG CTC TGG CGC CTG CCG GAG GAG
   E   S   A   A   R   P   A   A   G   P   A   L   W   R   L   P   E   E

         119         128         137         146         155         164
   CTG CTG CTG CTC ATC TGC TCC TAC CTG GAC ATG CGG GCC CTC GGC CGC CTG GCC
   L   L   L   L   I   C   S   Y   L   D   M   R   A   L   G   R   L   A

         173         182         191         200         209         218
   CAG GTG TGC CGC TGG CTG CGG CGC TTC ACC AGC TGC GAT CTG CTC TGG CGC CGG
   Q   V   C   R   W   L   R   R   F   T   S   C   D   L   L   W   R   R

         227         236         245         254         263         272
   ATA GCC CGG GCC TCG CTC AAC TCC GGC TTC ACG CGG CTC GGC ACC GAC CTG ATG
   I   A   R   A   S   L   N   S   G   F   T   R   L   G   T   D   L   M

         281         290         299         308         317         326
   ACC AGT GTC CCA GTG AAG GAA CGA GTG AAG GTG TCT CAG AAC TGG AGA CTG GGG
   T   S   V   P   V   K   E   R   V   K   V   S   Q   N   W   R   L   G

         335         344         353         362         371         380
   CGC TGC CGA GAG GGG ATT CTG CTG AAG TGG AGA TGC AGT CAG ATG CCC TGG ATG
   R   C   R   E   G   I   L   L   K   W   R   C   S   Q   M   P   W   M
```

FIGURE 2A

```
                389                 398                 407                 416                 425                 434
CAG CTA GAG GAT GAT TCT CTG TAC ATA TCC CAG GCT AAT TTC ATC CTG GCC TAC
Q   L   E   D   D   S   L   Y   I   S   Q   A   N   F   I   L   A   Y

                443                 452                 461                 470                 479                 488
CAG TTC CGT CCA GAT GGT GCC AGC TTG AAT CGT CGG CCT CTG GGA GTC TTT GCT
Q   F   R   P   D   G   A   S   L   N   R   R   P   L   G   V   F   A

                497                 506                 515                 524                 533                 542
GGG CAT GAT GAG GAC GTT TGC CAC TTT GTG CTG GCC AAC TCG CAT ATT GTT AGT
G   H   D   E   D   V   C   H   F   V   L   A   N   S   H   I   V   S

                551                 560                 569                 578                 587                 596
GCA GGA GGG GAT GGG AAG ATT GGC ATT CAT AAG ATT CAC AGC ACC TTC ACT GTC
A   G   G   D   G   K   I   G   I   H   K   I   H   S   T   F   T   V

                605                 614                 623                 632                 641                 650
AAG TAC TCG GCT CAT GAA CAG GAG GTG AAC TGT GTG GAT TGC AAA GGG GGC ATC
K   Y   S   A   H   E   Q   E   V   N   C   V   D   C   K   G   G   I

                659                 668                 677                 686                 695                 704
ATT GTG AGT GGC TCC AGG GAC AGG ACG GCC AAG GTG TGG CCT TTG GCC TCA GGC
I   V   S   G   S   R   D   R   T   A   K   V   W   P   L   A   S   G

                713                 722                 731                 740                 749                 758
CGG CTG GGG CAG TGC TTA CAC ACC ATC CAG ACT GAA GAC CGA GTC TGG TCC ATT
R   L   G   Q   C   L   H   T   I   Q   T   E   D   R   V   W   S   I
```

FIGURE 2B

```
                767                 776                 785                 794                 803                 812
GCT ATC AGC CCA TTA CTC AGC TCT TTT GTG ACA GGG ACG GCT TGT TGC GGG CAC
A   I   S   P   L   L   S   S   F   V   T   G   T   A   C   C   G   H

                821                 830                 839                 848                 857                 866
TTC TCA CCC CTG AGA ATC TGG GAC CTC AAC AGT GGG CAG CTG ATG ACA CAC TTG
F   S   P   L   R   I   W   D   L   N   S   G   Q   L   M   T   H   L

                875                 884                 893                 902                 911                 920
GGC AGT GAC TTT CCC CCA GGG GCT GGG GTG CTG GAT GTC ATG TAT GAG TCC CCT
G   S   D   F   P   P   G   A   G   V   L   D   V   M   Y   E   S   P

                929                 938                 947                 956                 965                 974
TTC ACA CTG CTG TCC TGT GGC TAT GAC ACC TAT GTT CGC TAC TGG GAC CTC CGC
F   T   L   L   S   C   G   Y   D   T   Y   V   R   Y   W   D   L   R

                983                 992                1001                1010                1019                1028
ACC AGC GTC CGG AAA TGT GTC ATG GAG TGG GAG GAG CCC CAC GAC AGC ACC CTG
T   S   V   R   K   C   V   M   E   W   E   E   P   H   D   S   T   L

               1037                1046                1055                1064                1073                1082
TAC TGC CTG CAG ACA GAT GGC AAC CAC CTG CTG GCC ACA GGT TCC TCC TAC TAC
Y   C   L   Q   T   D   G   N   H   L   L   A   T   G   S   S   Y   Y

               1091                1100                1109                1118                1127                1136
GGT GTT GTA CGG CTG TGG GAC CGG CGT CAA AGG GCC TGC CTG CAC GCC TTC CCG
G   V   V   R   L   W   D   R   R   Q   R   A   C   L   H   A   F   P
```

FIGURE 2C

```
       1145        1154        1163        1172        1181        1190
CTG ACG TCG ACT CCC CTC AGC AGC CCT GTG TAC TGC CTG CGT CTC ACC ACC AAG
L   T   S   T   P   L   S   S   P   V   Y   C   L   R   L   T   T   K

       1199        1208        1217        1226        1235        1244
CAT CTC TAT GCT GCC CTG TCT TAC AAC CTC CAC GTC CTG GAT TTT CAA AAC CCA
H   L   Y   A   A   L   S   Y   N   L   H   V   L   D   F   Q   N   P

       1253        1262        1271        1280        1289        1298
TGA CCG TCA GGG CCA CCC CTG CCT CTG GGC CAG GGA AAC CAG CTA CTC AGG GAC

       1307        1316        1325        1334        1343        1352
TTC TCT TGC CTG GAG GGT GCA GTG ATA GCT CCT CCT CAC TGC CCC ACT GTG CTC

       1361        1370        1379        1388        1397        1406
CTG GGC CTG TGA CCC CAG TGC TCA GGC ACC TTG CAC TAG AGG CTT CTG ACT CCT

       1415        1424        1433        1442        1451        1460
GGG ACT TTG GAG CTT ACC AGA GAT GCA GTC CCT CCC AGG AAC CTG TTG GAG AGG

       1469        1478        1487        1496        1505        1514
CAG GAC CTG CTG CTT TAG AGT GCG GCT GAA CCC GGG CCT TGC GTC CCT GTT TGG

       1523        1532        1541        1550        1559        1568
CCA GAG CAA GGA TCT GGC CTG GAG AGG CCC ATC CTA TAC CCC TTA TTA GAG CCA

       1577        1586        1595        1604        1613        1622
TGA CAG CCT ACA GAG TGA GGT GAG GTG CTC CCA CCT TCC CAG ATG GTT CCT TTC
```

FIGURE 2D

```
1631                1640                1649                1658                1667                1676
TGC CCC TTC CTG GAA GGA AAG GTG AGG CTG CCA ATA GCC TCC TGG CAC CAG CCA

1685                1694                1703                1712                1721                1730
GAC CTC ACC CTT GAC CAA CCT CTC GGG GCT GGG GGT TCA TTC CTG GGG CAC TGT

1739                1748                1757                1766                1775                1784
GGC CTG GTT TTG CTT TGA AAC CAA GAA AGA GCA AAG GGA ACC CAG CAG TTC TGA

1793                1802                1811                1820                1829                1838
GTG AGT TCT GAG CCA GCC CTA CCT CAG GCT GGC TGT TGA GAC ATG CTA CAA TTT

1847                1856                1865                1874                1883
TCA TTT TTG TAA AAA TAA AGC TTG ATT GTT CAC AGA AAA AAA AAA AAA AA 3'
```

FIGURE 2E

```
                9          18          27          36          45          54
5' AGC TGT GAG AGT GTG CAG TCG CGT TCC TGC TGT CCG GAC ACT TTT TTC CTC TAC

               63          72          81          90          99         108
   TGA GAC TCA TCT GGT AGA TCC GCA GGC CAG TCC TCC CAG GGG CTG AAG TTG TGA

              117         126         135         144         153         162
   AAT ATG GGT TTT CTA AGA AGA TTA ATC TAT CGG CGT AGA CCA ATG ATC TAT GTA
       M   G   F   L   R   R   L   I   Y   R   R   R   P   M   I   Y   V

              171         180         189         198         207         216
   GAA TCT TCT GAG GAG TCC AGT GAT GAG CAA CCT GAC GAA GTG GAA TCA CCA ACT
   E   S   S   E   E   S   S   D   E   Q   P   D   E   V   E   S   P   T

              225         234         243         252         261         270
   CAA AGT CAG GAT TCT ACA CCT GCT GAA GAG AGA GAG GAT GAG GGA GCA TCT GCA
   Q   S   Q   D   S   T   P   A   E   E   R   E   D   E   G   A   S   A

              279         288         297         306         315         324
   GCT CAA GGG CAG GAG CCT GAA GCT GAT AGC CAG GAA CTG GTT CAG CCA AAG ACT
   A   Q   G   Q   E   P   E   A   D   S   Q   E   L   V   Q   P   K   T

              333         342         351         360         369         378
   GGG TGT GAG CTT GGA GAT GGT CCT GAT ACC AAG AGG GTG TGC CTG CGA AAT GAA
   G   C   E   L   G   D   G   P   D   T   K   R   V   C   L   R   N   E
```

FIGURE 3A

```
                387                 396                 405                 414                 423                 432
GAG CAG ATG AAA CTG CCC GCA GAA GGG CCA GAG CCT GAA GCG GAT AGC CAG GAA
E   Q   M   K   L   P   A   E   G   P   E   P   E   A   D   S   Q   E

                441                 450                 459                 468                 477                 486
CAG GTT CAC CCG AAG ACT GGG TGT GAG CGC GGA GAT GGT CCT GAT GTC CAG GAG
Q   V   H   P   K   T   G   C   E   R   G   D   G   P   D   V   Q   E

                495                 504                 513                 522                 531                 540
TTG GGC CTG CCA AAT CCA GAG GAG GTG AAA ACA CCT GAG GAA GAT GAA GGG CAA
L   G   L   P   N   P   E   E   V   K   T   P   E   E   D   E   G   Q

                549                 558                 567                 576                 585                 594
TCA CAG CCT TAA AAG AAG ACA CGC TGA AAT GGT TCA GGC TGC TCC TGT GTT GGA
S   Q   P

                603                 612                 621                 630                 639                 648
AAT TTG ACC ATT AAA ATT CTC CCA ATA AAG CTT TAC AGC CTT CTG CAA AAA AAA

AA 3'
```

FIGURE 3B

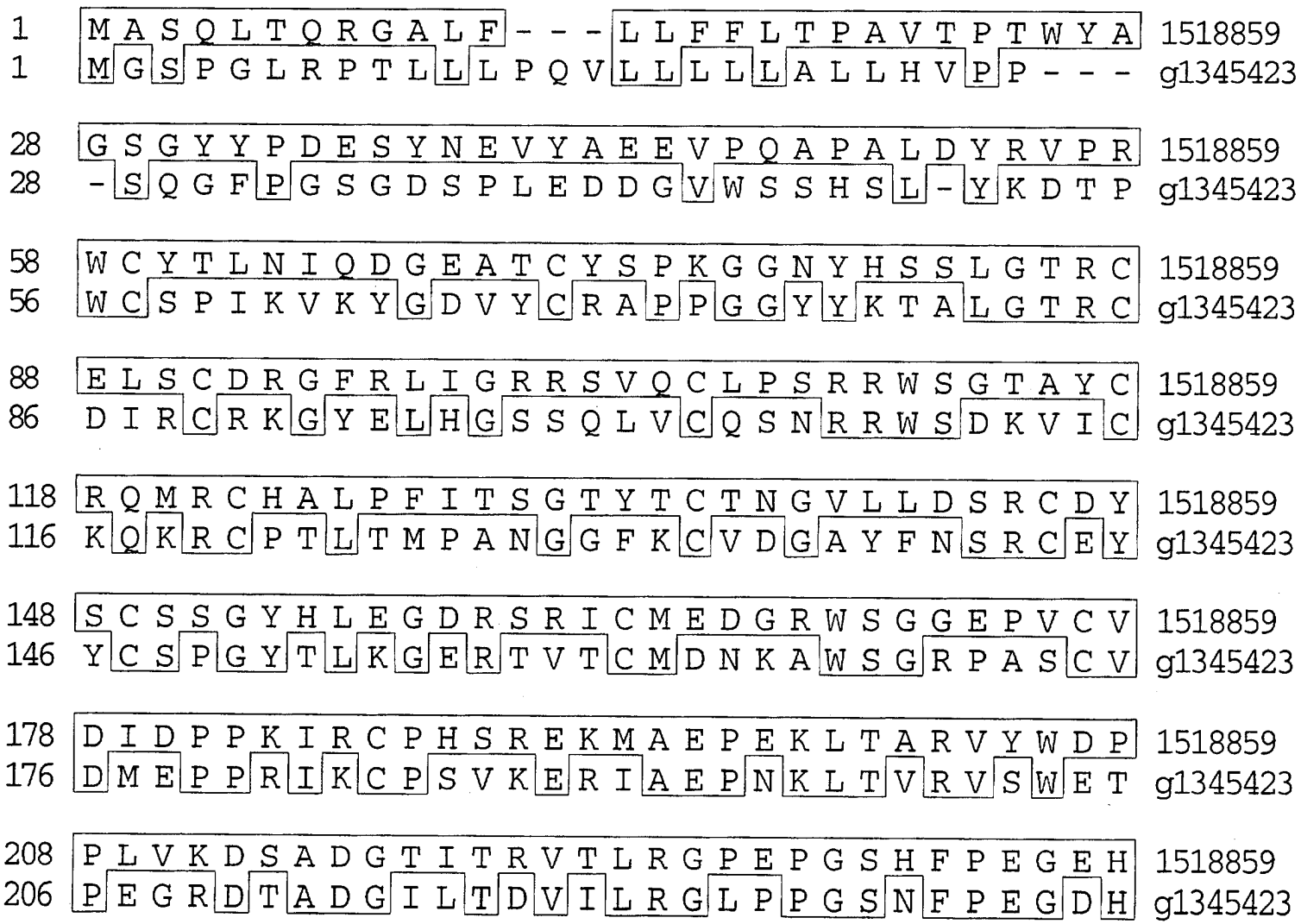
1 MASQLTQRGALF---LLFFLTPAVTPTWYA 1518859
1 MGSPGLRPTLLLPQVLLLLLALLHVPP--- g1345423
28 GSGYYPDESYNEVYAEEVPQAPALDYRVPR 1518859
28 -SQGFPGSGDSPLEDDGVWSSHSL-YKDTP g1345423
58 WCYTLNIQDGEATCYSPKGGNYHSSLGTRC 1518859
56 WCSPIKVKYGDVYCRAPPGGYYKTALGTRC g1345423
88 ELSCDRGFRLIGRRSVQCLPSRRWSGTAYC 1518859
86 DIRCRKGYELHGSSQLVCQSNRRWSDKVIC g1345423
118 RQMRCHALPFITSGTYTCTNGVLLDSRCDY 1518859
116 KQKRCPTLTMPANGGFKCVDGAYFNSRCEY g1345423
148 SCSSGYHLEGDRSRICMEDGRWSGGEPVCV 1518859
146 YCSPGYTLKGERTVTCMDNKAWSGRPASCV g1345423
178 DIDPPKIRCPHSREKMAEPEKLTARVYWDP 1518859
176 DMEPPRIKCPSVKERIAEPNKLTVRVSWET g1345423
208 PLVKDSADGTITRVTLRGPEPGSHFPEGEH 1518859
206 PEGRDTADGILTDVILRGLPPGSNFPEGDH g1345423

FIGURE 4A

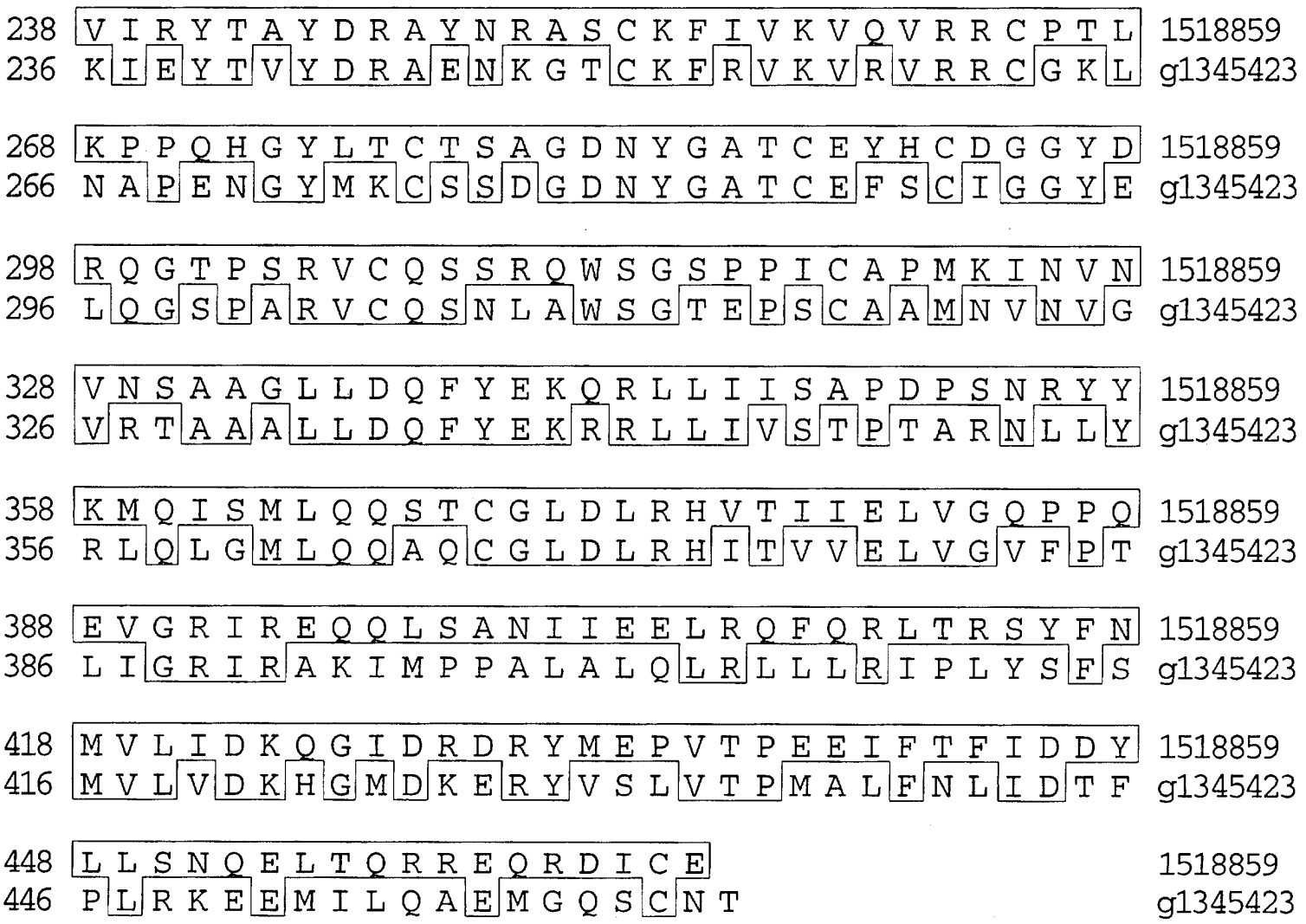
238 V I R Y T A Y D R A Y N R A S C K F I V K V Q V R R C P T L 1518859
236 K I E Y T V Y D R A E N K G T C K F R V K V R V R R C G K L g1345423
268 K P P Q H G Y L T C T S A G D N Y G A T C E Y H C D G G Y D 1518859
266 N A P E N G Y M K C S S D G D N Y G A T C E F S C I G G Y E g1345423
298 R Q G T P S R V C Q S S R Q W S G S P P I C A P M K I N V N 1518859
296 L Q G S P A R V C Q S N L A W S G T E P S C A A M N V N V G g1345423
328 V N S A A G L L D Q F Y E K Q R L L I I S A P D P S N R Y Y 1518859
326 V R T A A A L L D Q F Y E K R R L L I V S T P T A R N L L Y g1345423
358 K M Q I S M L Q Q S T C G L D L R H V T I I E L V G Q P P Q 1518859
356 R L Q L G M L Q Q A Q C G L D L R H I T V V E L V G V F P T g1345423
388 E V G R I R E Q Q L S A N I I E E L R Q F Q R L T R S Y F N 1518859
386 L I G R I R A K I M P P A L A L Q L R L L L R I P L Y S F S g1345423
418 M V L I D K Q G I D R D R Y M E P V T P E E I F T F I D D Y 1518859
416 M V L V D K H G M D K E R Y V S L V T P M A L F N L I D T F g1345423
448 L L S N Q E L T Q R R E Q R D I C E 1518859
446 P L R K E E M I L Q A E M G Q S C N T g1345423

FIGURE 4B

```
1    M - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   2616269
1    R G G S E G R G R G R E K R A R G A R R K R K Q G G R E A R   g487348

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   2616269
31   A A D G E G G S G P G A E A G A R T R P R E E A E G G G S V   g487348

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   2616269
61   E E G A R G I I K G D E G S V G A G K E A Q G R K Y G K E E   g487348

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   2616269
91   W R V R A R R R E G A R P G R V Q G Q G G Q V W A Y I P G T   g487348

2    - - - - - - - - - - - - - - A A R E S A A R P A A G P A L W   2616269
121  G A A M A A A A R E E E E E A A R E S A A C P A A G P A L W   g487348

18   R L P E E L L L L I C S Y L D M R A L G R L A Q V C R W L R   2616269
151  R L P E V L L L H M C S Y L D M R A L G R L A Q V Y R W L W   g487348

48   R F T S C D L L W R R I A R A S L N S G F T R L G T D L M T   2616269
181  H F T N C D L L R R Q I A W A S L N S G F T R L G T N L M T   g487348

78   S V P V K E R V K V S Q N W R L G R C R E G I L L K W R C S   2616269
211  S V P V K - - - - V S Q N W I V G C C R E G I L L K W R C S   g487348
```

FIGURE 5A

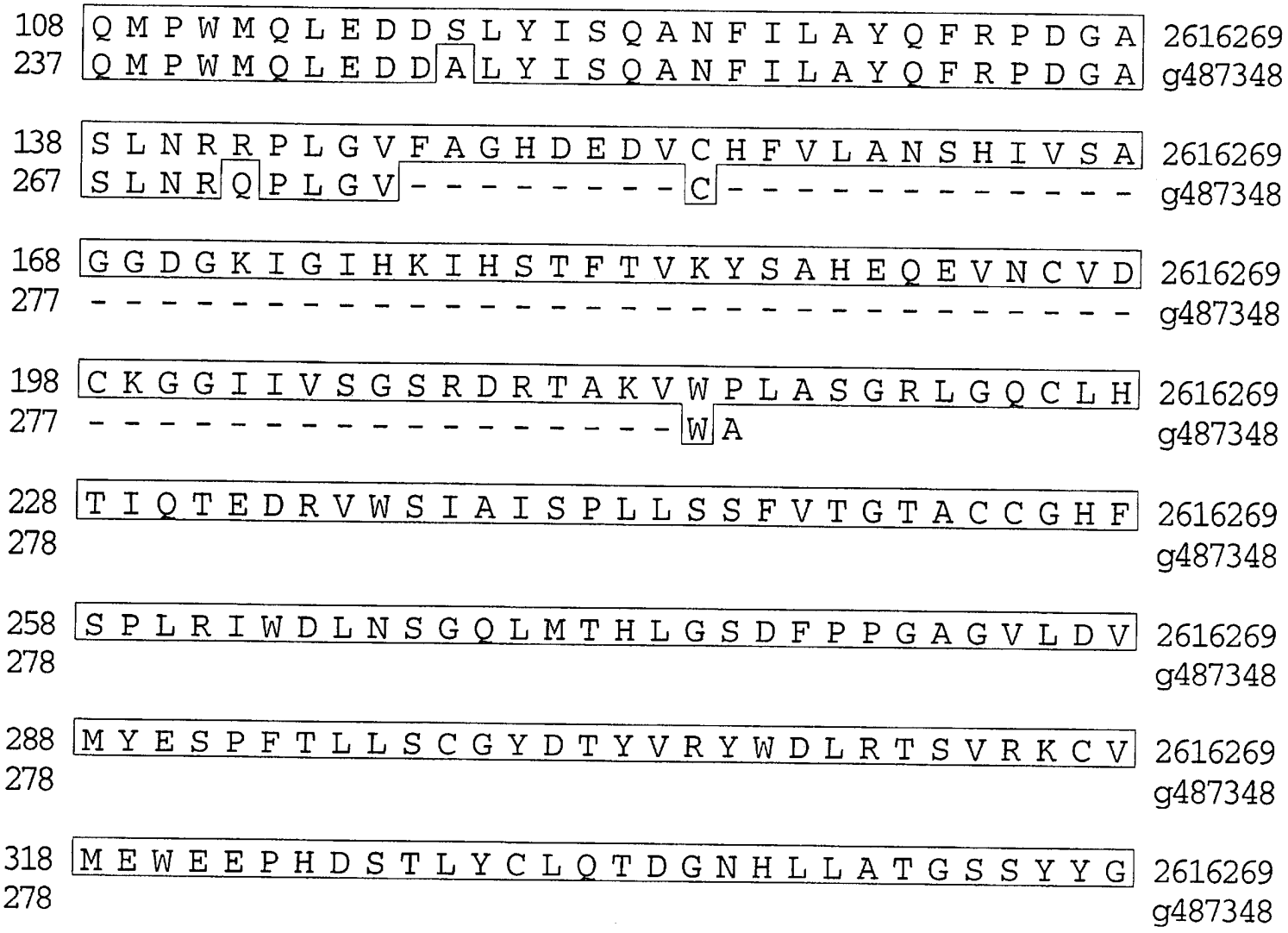
108 QMPWMQLEDDSLYISQANFILAYQFRPDGA 2616269
237 QMPWMQLEDDALYISQANFILAYQFRPDGA g487348
138 SLNRRPLGVFAGHDEDVCHFVLANSHIVSA 2616269
267 SLNRQPLGV----------C---------- g487348
168 GGDGKIGIHKIHSTFTVKYSAHEQEVNCVD 2616269
277 ---------------------------- g487348
198 CKGGIIVSGSRDRTAKVWPLASGRLGQCLH 2616269
277 ----------------------W A g487348
228 TIQTEDRVWSIAISPLLSSFVTGTACCGHF 2616269
278 g487348
258 SPLRIWDLNSGQLMTHLGSDFPPGAGVLDV 2616269
278 g487348
288 MYESPFTLLSCGYDTYVRYWDLRTSVRKCV 2616269
278 g487348
318 MEWEEPHDSTLYCLQTDGNHLLATGSSYYG 2616269
278 g487348

FIGURE 5B

348 V V R L W D R R Q R A C L H A F P L T S T P L S S P V Y C L 2616269

278 g487348

378 R L T T K H L Y A A L S Y N L H V L D F Q N P 2616269

278 g487348

FIGURE 5C

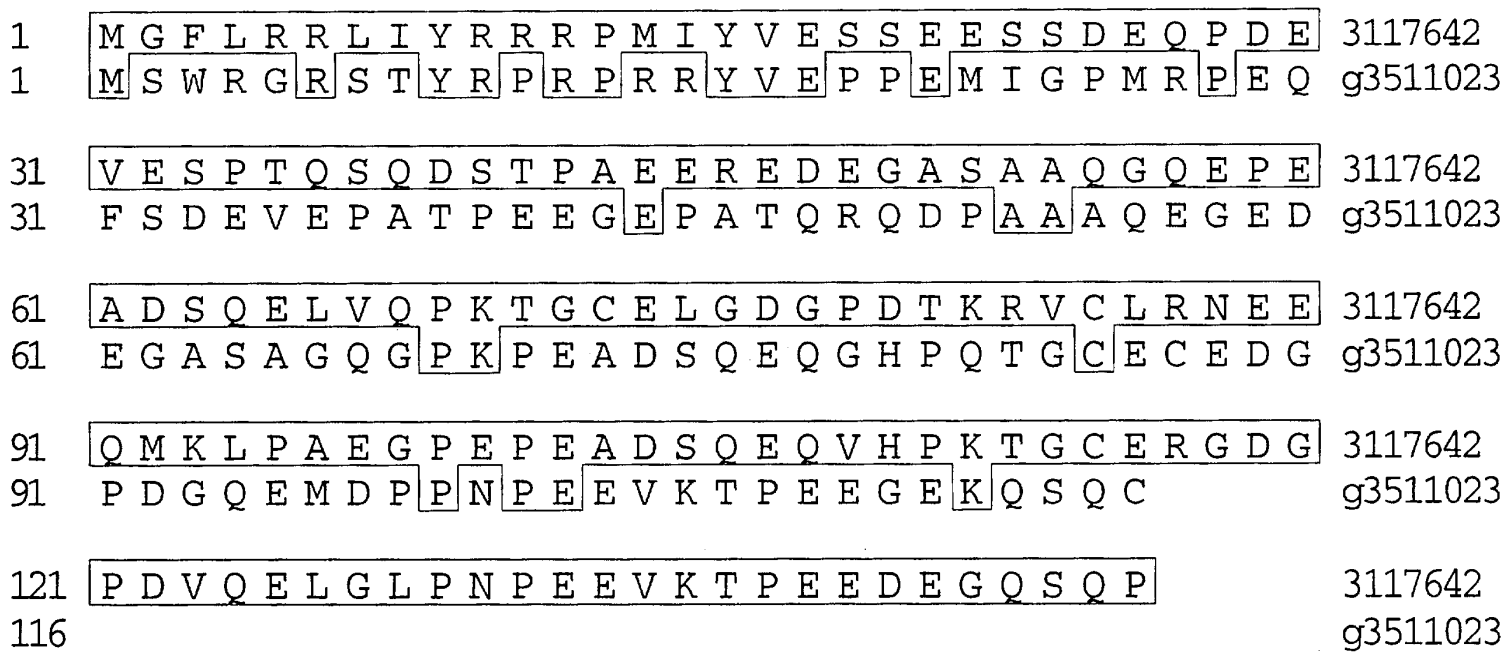
1 MGFLRRLIYRRRPMIYVESSEESSDEQPDE 3117642
1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQ g3511023
31 VESPTQSQDSTPAEEREDEGASAAQGQEPE 3117642
31 FSDEVEPATPEEGEPATQRQDPAAAQEGED g3511023
61 ADSQELVQPKTGCELGDGPDTKRVCLRNEE 3117642
61 EGASAGQGPKPEADSQEQGHPQTGCECEDG g3511023
91 QMKLPAEGPEPEADSQEQVHPKTGCERGDG 3117642
91 PDGQEMDPPNPEEVKTPEEGEKQSQC g3511023
121 PDVQELGLPNPEEVKTPEEDEGQSQP 3117642
116 g3511023

FIGURE 6

Docket No.: PF-0661 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

CANCER ASSOCIATED PROTEINS

the specification of which:

/___/ is attached hereto.

/ X / was filed on (to be assigned) as application Serial No. 09/889,617 and if this box contains an X /___/, was amended on _____.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US00/01565 on January 21, 2000, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on _____ 2001, and if this box contains an X /_/, was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Docket No.: PF-0661 USN

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	/_/ Yes /_/ No
_____	_____	_____	/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/183,027	January 22, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Matthew R. Kaser	Reg. No. 44,817
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. 47,139
P. Ben Wang	Reg. No. 41,420

13

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0661 USN

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

First Joint Inventor:	Full name:	Jennifer L. Hillman
	Signature:	Jennifer L Hillman
	Date:	September 21, 2001
	Citizenship:	United States
	Residence:	Mountain View, California
	P.O. Address:	230 Monroe Drive, #17 Mountain View, California 94040
Second Joint Inventor:	Full name:	Henry Yue
	Signature:	Henry Yue
	Date:	September 24, 2001
	Citizenship:	United States
	Residence:	Sunnyvale, California
	P.O. Address:	826 Lois Avenue Sunnyvale, California 94087

Docket No.: PF-0661 USN

Third Joint Inventor:

Full name: 3-00 Y. Tom Tang

Signature: Y. Tom Tang

Date: September 10, 2001

Citizenship: United States

Residence: San Jose, California

P.O. Address: 4230 Ranwick Court CA
San Jose, California 95118

Fourth Joint Inventor:

Full name: 4-00 Yalda Azimzai

Signature: Yalda Azimzai

Date: September 13, 2001

Citizenship: United States

Residence: Castro Valley, California

P.O. Address: 5518 Boulder Canyon Drive CA
Castro Valley, California 94552

[redacted]